REMARKS

Applicant has amended claims 16 and 23, and have cancelled claims 15 and 24-38, during prosecution of this patent application. Applicant is not conceding in this patent application that the subject matter encompassed by said amended and cancelled claims are not patentable over the art cited by the Examiner, since the claim amendments and cancellations are only for facilitating expeditious prosecution of this patent application. Applicant respectfully reserves the right to pursue the subject matter encompassed by said amended and cancelled claims, and to pursue other claims, in one or more continuations and/or divisional patent applications

Claim 16 has been rewritten in independent form and is otherwise essentially the same claim 16 as previously existed.

The Examiner rejected claims 15-38 under 35 U.S.C. § 102(e) as allegedly being anticipated by U.S. Patent No. 7,146,353 to Garg, Pankaj K. et al., (hereinafter "Garg").

Applicant respectfully traverses the § 102 rejections with the following arguments.

35 U.S.C. § 102

The Examiner rejected claims 15-38 under 35 U.S.C. § 102(e) as allegedly being anticipated by U.S. Patent No. 7,146,353 to Garg, Pankaj K. et al., (hereinafter "Garg").

Since claims 15 and 24-38 have been canceled, the rejection of claims 15 and 24-38 under 35 U.S.C. § 102(e) is moot.

Applicant respectfully contends that Garg does not anticipate claim 15, because Garg does not teach each and every feature of claim 16.

As a first example of why Garg does not anticipate claim 16, Garg does not teach the feature: "monitoring a processing time required for each application program to process the transaction received by each application server".

The Examiner argues that Garg, col. 4, lines 19-21, col. 7, lines 58-67 teaches the preceding feature of claim 16.

In response, Applicant notes that Garg, col. 4, lines 19-21 discusses ***mean resonse time*** of a request and is totally silent as to ***the processing time required for each application program*** to process the transaction received by each application server. In other words, Garg teaches mean times and does not teach individual times. In particular, Grag, col. 5, lines 30-34 and 55-59 teaches a sum of average service demands in terms of various parameters, none of which being a processing time or a response time of each application program.

In further response, Applicant asserts that Garg, col. 7, lines 58-67 is totally silent as to "monitoring a processing time required for each application program to process the transaction received by each application server".

Therefore, Garg does not teach the preceding feature of claim 16.

As a second example of why Garg does not anticipate claim 16, Garg does not teach the feature: "wherein said detecting the bottleneck relating to usage of at least one resource comprises identifying the at least one resource, and wherein said identifying the at least one resource comprises independently identifying each resource of the at least one resource as being ... **said resource of at least one application server of the N application servers if I is at least 1 and does not exceed M and if a processing time for processing another type of transaction by any application server of the N application servers is not within the predesignated permissible processing time range**" (emphasis added).

The Examiner argues that "Garg teaches ... said resource of at least one application server of the N application servers if I is at least 1 and does not exceed M and if a processing time for processing another type of transaction by any application server of the N application servers is not within the predesignated permissible processing time range (as stated in col. 5, lines 65-67, col. 6, lines 1-13, ***mathematical optimization*** model is ***formulated*** to ***find*** the ***optimal number*** *of* ***servers*** at each of the ***tiers.*** The ***decision variables*** on which ***optimization*** are ***performed*** *is* the ***number of servers*** at ***each tier*** in the multi-tiered system. The objective function is the weighted sum of the number of servers at each tier, where the weights are the "costs" per server. The number of servers at each tier is constrained to be an integer greater than or equal to one. ***Optimization model*** has constraint E[R].Itoreq.SLA.sub.R, where SLA.sub.R is the ***response time limit*** such as 1 second ***required*** by the ***Service Level Agreement*** SLA".

In response, Applicant respectfully contends that the preceding argument by the Examiner has not addressed the preceding feature of claim 16 which recites that if a particular condition is satisfied, then the detected bottleneck relates to usage of resources identified as application servers. The particular condition that must be satisfied is: "I is at least 1 and does not exceed M and if a processing time for processing another type of transaction by any application server of the N application servers is not within the predesignated permissible processing time range".

Applicant asserts that Garg, col. 5, lines 65-67, col. 6, lines 1-13 does not teach that the detected bottleneck relates to usage of resources identified as application servers if the preceding particular condition is satisfied. Moreover, the preceding discussion by the Examiner is irrelevant and does not argue that the detected bottleneck relates to usage of resources identified as application servers if the preceding particular condition is satisfied.

Therefore, Garg does not teach the preceding feature of claim 16

As a third example of why Garg does not anticipate claim 16, Garg does not teach the feature: "wherein said detecting the bottleneck relating to usage of at least one resource comprises identifying the at least one resource, and wherein said identifying the at least one resource comprises independently identifying each resource of the at least one resource as being ... **said resource related to input to the transaction if I is at least 1 and does not exceed M and if a processing time for processing another type of transaction by any application server of the N application servers is within the predesignated permissible processing time range**"(emphasis added).

The Examiner argues that "Garg teaches ... said resource related to input to the transaction if I is at least 1 and does not exceed M and if a processing time for processing another type of transaction by any application server of the N application servers is within the predesignated permissible processing time range (as stated in col. 6, lines 14-19, resulting ***mathematical optimization model*** has *a* ***linear objective*** function but a ***non-linear,*** inequality-type constraint with integer-valued ***decision variables. A concavity*** property of the ***average response time*** E[R] ***function*** *is* used with respect to the ***decision variables*** in formulating an ***efficient*** bounding ***procedure)***".

In response, Applicant respectfully contends that the preceding argument by the Examiner has not addressed the preceding feature of claim 16 which recites that if the preceding particular condition is satisfied, then the detected bottleneck relates to usage of resources identified as input to the transaction. The particular condition that must be satisfied is: "I is at least 1 and does not exceed M and if a processing time for processing another type of transaction by any application server of the N application servers is within the predesignated permissible processing time range".

Applicant asserts that Garg, col. 6, lines 14-19 does not teach that the detected bottleneck relates to usage of resources identified as input to the transaction if the preceding particular condition is satisfied. Moreover, the preceding discussion by the Examiner is irrelevant and does not argue that the detected bottleneck relates to usage of resources identified as input to the transaction if the preceding particular condition is satisfied.

Therefore, Garg does not teach the preceding feature of claim 16

As a fourth example of why Garg does not anticipate claim 16, Garg does not teach the feature: "wherein said detecting the bottleneck relating to usage of at least one resource comprises identifying the at least one resource, and wherein said identifying the at least one resource comprises independently identifying each resource of the at least one resource as being ... **said resource of the database server if I exceeds M and if a processing time for processing another type of transaction by any application server of the N application servers is not within the predesignated permissible processing time range**"(emphasis added).

The Examiner argues that "Garg teaches ... said resource of the database server if I exceeds M and if a processing time for processing another type of transaction by any application server of the N application servers is not within the predesignated permissible processing time range (as stated in col. 6, lines 19-24, bounding procedure ignores the integer-value requirements on the ***decision variables*** **and** ***solves*** the ***2-tiered problem.*** The ***solution*** *is* then ***rounded*** to ***integer values.*** Then the ***3-tiered problem*** *is* ***solved*** using the ***solution*** to the **2-*tiered problem*** and recursively to the general ***n-tiered problem)***".

In response, Applicant respectfully contends that the preceding argument by the Examiner has not addressed the preceding feature of claim 16 which recites that if a particular condition is satisfied, then the detected bottleneck relates to usage of resources identified as the database server. The particular condition that must be satisfied is: "I exceeds M and if a processing time for processing another type of transaction by any application server of the N application servers is not within the predesignated permissible processing time range".

Applicant asserts that Garg, col. 6, lines 19-24 does not teach that the detected bottleneck relates to usage of resources identified as the database server if the preceding particular

condition is satisfied. Moreover, the preceding discussion by the Examiner is irrelevant and does not argue that that the detected bottleneck relates to usage of resources identified as the database server if the preceding particular condition is satisfied.

Therefore, Garg does not teach the preceding feature of claim 16

As a fifth example of why Garg does not anticipate claim 16, Garg does not teach the feature: "wherein said detecting the bottleneck relating to usage of at least one resource comprises identifying the at least one resource, and wherein said identifying the at least one resource comprises independently identifying each resource of the at least one resource as being ... **said resource related to the transaction if I exceeds M and if a processing time for processing another type of transaction by any application server of the N application servers is within the predesignated permissible processing time range**"(emphasis added).

The Examiner argues that "Garg teaches ... said resource related to the transaction if I exceeds M and if a processing time for processing another type of transaction by any application server of the N application servers is within the predesignated permissible processing time range (as stated in col.6, lines 25-32, Once the ***server requirements*** have been ***estimated*** and ***optimized,*** an ***assignment*** of ***applications to servers*** may be ***determined*** *as a* ***function*** *of* the ***optimal server requirements*** predicted in such a way ***communications delays*** are ***minimized*** **and** ***bandwidth capacity constraints*** are ***satisfied*** (step 212). The ***bandwidth capacity constraints*** are the ***actual bandwidth*** of the physical **resources**".

In response, Applicant respectfully contends that the preceding argument by the Examiner has not addressed the preceding feature of claim 16 which recites that if a particular

condition is satisfied, then the detected bottleneck relates to resources identified as the transaction. The particular condition that must be satisfied is: "I exceeds M and if a processing time for processing another type of transaction by any application server of the N application servers is within the predesignated permissible processing time range".

Applicant asserts that Garg, col. 6, lines 25-32 does not teach that the detected bottleneck relates to usage of resources identified as the transaction if the preceding particular condition is satisfied. Moreover, the preceding discussion by the Examiner is irrelevant and does not argue that the detected bottleneck relates to usage of resources identified as the transaction if the preceding particular condition is satisfied.

Therefore, Garg does not teach the preceding feature of claim 16

Based on the preceding arguments, Applicants respectfully maintain that Garg does not anticipate claim 16, and that claim 16 is in condition for allowance. Since claims 17-23 depend from claim 15, Applicants contend that claims 17-23 are likewise in condition for allowance.

In addition with respect to claim 17, Garg does not teach the feature: "wherein the method further comprises monitoring processing loads imposed on: resources of the N application servers, resources of the database server, and resources related to the transaction, and wherein said identifying each resource of the at least one resource comprises determining from the monitored processing loads that a high load specific to each resource of the at least one resource is imposed on each resource of the at least one resource".

The Examiner argues: "**As to *Claims 17, 26 and 34,*** *Garg teaches method, Load Control Server and Computer readable medium, of claims 16, 25 and 33, wherein the method further comprises monitoring processing loads imposed on* (as stated in col. 6, lines 65-67, col. lines 18-20, FIG. 3 is a ***functional*** block diagram of an example arrangement for ***gathering (monitoring) data*** to be used in ***analyzing resource requirements*** and ***allocations*** for ***applications*** hosted by a data center. The ***load balancer*** distributes ***transactions*** in a manner that ***minimizes response time*** **and** ***maximizes resource utilization):*** *resources of the N application servers, resources of the database server, and resources related to the transaction, and wherein on* (as stated in col. 7, lines 36-38, ***Collector*** *332* ***gathers instrumentation data*** pertaining to web ***transactions*** *as* the ***transactions are processed*** by each ***component from block*** *302* ***to database*** 316); said identifying each resource of the at least one resource comprises determining from the monitored processing loads that a high load specific to each resource of the at least one resource is imposed on each resource of the at least one resource (as stated in col. 7, lines 45-54, ***Analyzer optimizer*** block 342 analyzes the ***correlated*** instrumentation ***data, determines*** *a* ***desired configuration,*** and initiates ***reconfiguration*** *of* the ***load balancer*** *306, servers* and ***load balancers*** in the web server farm 308, and **servers** in ***application server farm*** 314 *as* may be ***desirable.*** In an example embodiment, the analyzer-optimizer uses a ***queuing model*** to ***estimate*** **and** ***optimize server requirements*** *of* the ***applications*** based on mix of ***transaction types,*** the ***volume of the different transaction types,*** and *a* ***level of service*** that the data center is ***expected to provide)***".

In response, Applicants respectfully contend that the preceding argument by the Examiner has not addressed the peceding feature of claim 17 and is irrelevant to the preceding

feature of claim 17. The Examiner has not provided any analysis that links the the Examiner's argument to the language of the preceding feature of claim 17. Therefore, the Examiner's argument is not persuasive.

Therefore, Garg does not anticipate claim 17.

In addition with respect to claim 18, Garg does not teach the feature: "wherein said determining that a high load is imposed on each resource of the at least one resource comprises determining that a predesignated detection condition has occurred for each resource of the at least one resource a predesignated number of times, and wherein the predesignated detection condition is that a predesignated usage parameter specific to each resource of the at least one resource is in a predesignated load range".

The Examiner argues: "As **to Claims 18, and 27,** Garg teaches method, Load Control Server and Computer readable medium, of claims 17, and 26, wherein said determining that a high load is imposed on each resource of the at least one resource comprises (as stated in col. 4, lines 8-10, Based on the ***user classifications, workload mix,*** **and** ***workload levels*** the ***load balancing policies*** *may* be **adjusted):** determining that a predesignated detection condition has occurred for each resource of the at least one resource a predesignated number of times, and wherein the *predesignated detection condition is that a predesignated usage parameter specific to each resource of the at least one resource is in a predesignated load range (as* stated in col. 4, lines 34-43, The ***benefits*** of a ***sophisticated distribution policy*** based on ***user and URI classification*** may be sufficient to ***merit reassigning a session*** after the ***request*** in the ***session*** have been observed over some ***period of time.*** *This* may provide a more ***accurate estimate*** of the

sizes of subsequent ***requests*** in the session, if shortly into a ***session*** it is ***determined*** that the ***session*** *is* ***driven*** by a ***robot*** that will *issue* ***one type*** of ***request*** *a* **large** ***number of times,*** it might be worthwhile to ***reassign*** the ***session*** to a ***server dedicated*** to those ***types*** of ***requests)***".

In response, Applicants respectfully contend that the preceding argument by the Examiner has not addressed the peceding feature of claim 18 and is irrelevant to the preceding feature of claim 18. The Examiner has not provided any analysis that links the the Examiner's argument to the language of the preceding feature of claim 18. Therefore, the Examiner's argument is not persuasive.

Therefore, Garg does not anticipate claim 18.

In addition with respect to claim 19, Garg does not teach the feature: "wherein said removing the detected bottleneck comprises eliminating the high load imposed on each resource of the at least one resource".

The Examiner argues: "**As to Claims 19, 28 and 35,** Garg teaches method, Load Control Server and Computer readable medium, of claims 17, 26, and 34, wherein said removing the detected bottleneck comprises eliminating the high load imposed on each resource of the at least one (as stated in col. 7, lines 45-54, ***Analyzer-optimizer*** block 342 ***analyzes*** the ***correlated*** instrumentation ***data, determines*** *a* ***desired configuration,*** and initiates ***reconfiguration*** of the ***load balancer*** 306, ***servers*** and ***load balancers*** in the web server farm 308, and servers in ***application server farm*** 314 as may be ***desirable.*** The ***analyzer optimizer*** *uses* a ***queuing model*** to ***estimate*** **and** ***optimize server requirements*** *of* the ***applications*** based on *mix* of ***transaction***

types, the ***volume (load) of the different transaction types,*** and ***a level of service*** that the data center is expected to provide)."

In response, Applicants respectfully contend that the preceding argument by the Examiner has not addressed the peceding feature of claim 19 and is irrelevant to the preceding feature of claim 19. The Examiner has not provided any analysis that links the the Examiner's argument to the language of the preceding feature of claim 19. Therefore, the Examiner's argument is not persuasive.

Therefore, Garg does not anticipate claim 19.

In addition with respect to claim 20, Garg does not teach the feature: "wherein said eliminating comprises executing in a predesignated sequence specific to each resource of the at least one resource as many of one or more predesignated load control processes as is necessary to eliminate the high load imposed on each resource of the at least one resource".

The Examiner argues: "As **to Claims 20, 29 and 36,** Garg teaches method, Load Control Server and Computer readable medium, of claims 19, 28, and 35, wherein said eliminating comprises executing in a predesignated sequence specific to each resource of the at least one resource as many of one or more predesignated load control processes as is necessary to eliminate the high load imposed on each resource of the at least one resource (as stated in col. 4, lines 19-33, ***objective*** of a ***load distribution policy*** *is* to minimize some criterion such as the mean response time of a ***request*** by implementing several known load distribution policies. The ***load*** may be ***distributed*** based on a ***round-robin, random, least-work-remaining*** or ***size-based policy. Sessions*** are considered in ***load balancing.*** **A** ***session*** *is a* ***sequence*** of ***related*** Web

requests. In the example policy, the assignment for routing is performed once per session. To ***implement a minimizing-variance*** aspect of the ***size-based policy,*** at the initial Web ***request*** of a ***session*** an ***estimation*** of the **size** of ***subsequent requests*** *is* made. ***Sessions*** comprising mostly ***small requests*** may be ***assigned*** to ***different servers*** from those ***comprising*** mostly ***large requests).***"

In response, Applicants respectfully contend that the preceding argument by the Examiner has not addressed the peceding feature of claim 20 and is irrelevant to the preceding feature of claim 20. The Examiner has not provided any analysis that links the the Examiner's argument to the language of the preceding feature of claim 20. Therefore, the Examiner's argument is not persuasive.

Therefore, Garg does not anticipate claim 20.

In addition with respect to claim 21, Garg does not teach the feature: "wherein a first resource of the at least one resource is a resource of a first application server of the N application servers, wherein said executing the predesignated sequence specific to the first resource comprises reducing an application program multiplicity of the first application server, and wherein said application program multiplicity on the first application server is defined as a maximum number of application programs to be executed concurrently on the first application server with respect to a plurality of transactions of the same type that were received by the first application server at the same time".

The Examiner argues: "**As to *Claims 21, 30 and 37,*** *Garg teaches method, Load Control Server and Computer readable medium, of claims 20, 29, and 36, wherein a first resource of the*

at least one resource is a resource of a first application server of the N application servers, wherein said executing the predesignated sequence specific to the first resource comprises (as stated in col. 12, lines 37-39, lines 65-67, ***mathematical*** formulation of the ***resource allocation problem*** (RAP) *is* described as follows. The ***application architecture*** requirements are represented by the following ***parameters.*** The number **of servers** to be ***allocated*** to *tier I* is defined by **N.sub.l)...** reducing an application program multiplicity of the first application server, and wherein said application program multiplicity on the first application server is defined as a maximum number of application programs to be executed concurrently on the first application server with respect to a plurality of transactions of the same type that were received by the first application server at the same time (as stated in col. 13, lines 1-39, col. 15, lines 62-67, The ***maximum*** and ***minimum attribute*** requirements are represented by two matrices VMAX and VMIN, where each element VMAX.sub.la and VMIN.sub.la represent the maximum and minimum ***level of attribute*** *a* for any ***server in tier*** *I.* The matrix T is defined to ***characterize the traffic*** pattern of the ***application,*** where the ***element T.sub.li*** represents the ***maximum amount*** of ***traffic*** going from each **server** ***in tier*** *Ito* each ***server in tier*** *i.* The numbers T.sub.01 and T.sub.10 represent the Internet traffic coming into and going out of each server in tier I. Using these traffic parameters, the ***total amount*** of incoming and outgoing ***traffic*** at each ***server*** in different tiers may be ***calculated,*** denoted by *Tl.sub.l and TO.sub.l,* respectively. To reduce the ***number*** of ***binary variables x.sub.ls*** in the formulation, a ***feasibility matrix F*** *is* defined by Mixed Integer Programming problem, MIP2, to intelligently round the local optimal solution generated by the QP model. The MIP2 model ***defines*** the ***actual*** **servers** to ***allocate*** to the ***application.*** The ***decision variables*** are the ***same*** as those in the ***original problem P0).***"

In response, Applicants respectfully contend that the preceding argument by the Examiner has not addressed the peceding feature of claim 21 and is irrelevant to the preceding feature of claim 21. The Examiner has not provided any analysis that links the the Examiner's argument to the language of the preceding feature of claim 21. Therefore, the Examiner's argument is not persuasive.

Therefore, Garg does not anticipate claim 21.

In addition with respect to claim 22, Garg does not teach the feature: "wherein a first resource of the at least one resource is a resource of the database server, and wherein said executing the predesignated sequence specific to the first resource comprises reducing a priority level of a process for accessing the database".

The Examiner argues: "As **to Claims 22, 31 and 38,** Garg teaches method, Load Control Server and Computer readable medium, of claims 20, 29, and 36, wherein a first resource of the at least one resource is a resource of the database server, and wherein said executing the predesignated sequence specific to the first resource comprises reducing a priority level of a process for accessing the database (as stated in col. 1, lines 58-60, col. 6, lines *49-62, for each* application ***resource requirement*** *may* be determined *as a function of* the ***workload levels, service level*** metric associated with the application and ***subset*** **of resources** for each ***application.*** The adjusting ***load balancing policies,*** determining an ***allocation of resources,*** and ***automatically reconfiguring*** may be repeated as often as deemed necessary to achieve ***desired levels*** *of* ***performance*** **and** ***efficiency, reconfiguration*** tasks may include ***removing*** and installing application software, changing registry settings, editing of configuration files, and

running a command *to* start *the* application *software.* The *various* ***scripts*** and ***sequences*** *of* operations needed for ***reconfiguration*** will vary according to the ***type of server*** **and** ***characteristics*** of the ***application software).***"

In response, Applicants respectfully contend that the preceding argument by the Examiner has not addressed the peceding feature of claim 22 and is irrelevant to the preceding feature of claim 22. The Examiner has not provided any analysis that links the the Examiner's argument to the language of the preceding feature of claim 22. Therefore, the Examiner's argument is not persuasive.

Therefore, Garg does not anticipate claim 22.

In addition with respect to claim 23, Garg does not teach the feature: "wherein an upper limiting processing time of the predesignated permissible processing time range is one standard deviation higher than an average processing time per transaction processed during peak processing loads during a predesignated period of time".

The Examiner argues: "As **to Claim 23,** Garg teaches method, of claim 15, wherein an upper limiting processing time of the predesignated permissible processing time range is one standard deviation higher than an average processing time per transaction processed during peak processing loads during a predesignated period of time (as stated in col. 4, lines 34-37, col. 5, lines 60-64, The benefits of a sophisticated ***distribution policy based*** *on user* and URI classification *may* be *sufficient* to *merit* ***reassigning*** *a* ***session*** after the request in the session have been observed over some ***period of time*** and to approximate the ***average response time*** for a given number of servers at each tier, and an ***optimization process*** determines the minimum number of total

servers required for the ***application average response time*** to be within ***time range*** of ***specific SLA, service level agreement*** **and** ***average queuing time*** of the ***multi-tiered system*** then ***becomes*** the ***response time*** of the tiered system after ***adding*** to it some ***fixed "overhead" delays*** at non-bottleneck **resources** such as the ***fixed processing time*** **at** ***the load balancer).***"

In response, Applicants respectfully contend that the preceding argument by the Examiner has not addressed the peceding feature of claim 23 and is irrelevant to the preceding feature of claim 23. The Examiner has not provided any analysis that links the the Examiner's argument to the language of the preceding feature of claim 23. Therefore, the Examiner's argument is not persuasive.

Therefore, Garg does not anticipate claim 23.

CONCLUSION

Based on the preceding arguments, Applicants respectfully believe that all pending claims and the entire application meet the acceptance criteria for allowance and therefore request favorable action. If the Examiner believes that anything further would be helpful to place the application in better condition for allowance, Applicants invites the Examiner to contact Applicants' representative at the telephone number listed below. The Director is hereby authorized to charge and/or credit Deposit Account 09-0457 (IBM).

Date: June 18, 2008

/ Jack P. Friedman /
Jack P. Friedman
Registration No. 44,688

Customer No.: 30449
Schmeiser, Olsen & Watts
22 Century Hill Drive - Suite 302
Latham, New York 12110
Telephone (518) 220-1850
Facsimile (518) 220-1857
E-mail: jfriedman@iplawusa.com